



# Setting the grade standards of new GCSEs in England

SCORE's response to the Ofqual consultation

30 June 2014

## Introduction

1. SCORE is a partnership of organisations, which aims to improve science education in UK schools and colleges by supporting the development and implementation of effective education policy. The partnership is currently chaired by Professor Julia Buckingham and comprises the Association for Science Education, Institute of Physics, Royal Society, Royal Society of Chemistry and Society of Biology.
2. In summary:
  - The main purpose for the reform of GCSE qualifications should be a desire to improve educational outcomes for students by raising the quality of assessment. This is especially relevant now that all students must remain in compulsory education or training until 18. Ofqual's consultation document does not clearly define a purpose for examinations at key stage 4.
  - Ofqual heavily emphasises assessment at key stage 4 as a means by which to raise school performance through tightened accountability measures. These are primarily intended to be used in league table performance that can provide national and international comparisons. SCORE does not agree with this prioritisation of purposes.
  - Ofqual's assertion that its approach to standardisation is more reliable than others requires further clarification. SCORE supports a move to a more norm-referenced approach to standard setting in science.
  - SCORE is concerned that by making an explicit link between grade C and grade 4, the current emphasis on this grade as a 'pass' will undermine the achievement of the lower grades.
  - Ofqual makes no mention of the need to standardise across subjects for a long-term approach to grade-setting in future years.

## The purpose of GCSE qualifications

3. Now that students will remain in compulsory education up until the age of 18, student progress will be indicated by their leaving examinations at age 18. As such, SCORE does not feel that the reform of GCSE qualifications has gone far enough to clarify the purpose of examination at the age of 16, and resulting from this, does not present a clear case for assessment.
4. In 2013, Ofqual stated in its GCSE consultation that the 'proposed primary purposes of the reformed GCSEs will be to provide evidence of students' achievements against demanding and fulfilling content and a strong foundation for further academic and vocational study and for employment.'<sup>1</sup> SCORE responded to the Department for Education consultation on key stage 4 and the Ofqual consultation on GCSE reform emphasising that a reform of GCSEs presents the 'ideal opportunity to re-examine the purpose and nature of qualifications at key stage 4.'<sup>2</sup> In this 'Consultation on Setting the Grade Standards of new GCSEs in England' Ofqual makes no statement of purpose for assessment at key stage 4, or the grading structure itself.
5. Throughout the consultation document, grade-setting for accountability purposes is prioritised as a driver of reform. In doing so, Ofqual undermines the ultimate goal for education and assessment which is to be educationally useful to every student and to encourage and measure effective teaching and learning. Assessment must support high quality learning activities in schools at key stage 4. While the use of GCSE performance as a measure of school accountability might provide

---

<sup>1</sup> Ofqual, Consultation on GCSE reform, 2013 <http://comment.ofqual.gov.uk/gcse-reform-june-2013/>

<sup>2</sup> Both consultation responses are available on the SCORE website: <http://www.score-education.org/news/consultation-responses>

one purpose for GCSEs it should not take precedence over encouraging appropriate and positive educational outcomes.<sup>3</sup>

SCORE states that the purposes of GCSEs should be:

- To give every student an indication of their progress within the cohort as a reference point prior to the end of compulsory education at 18.
- In acting as a reference point, to provide students with a better understanding of their abilities in order to assist them in making decisions about further study and which courses to choose.
- To act as an achievement level guide in applications to universities, further education and employment (in the absence of other indicators, such as final AS- or A- level grades).
- To encourage effective teaching and learning through high quality and coherent curriculum and assessment development.

SCORE recommends that:

- Ofqual prioritises making a clear case for *why* performance standards should be set at this level alongside explaining how the standards will be set.
- Ofqual prioritises the main purpose of GCSEs as ensuring that new standards support good educational outcomes rather than emphasising the purpose of GCSEs as an accountability tool.

## Grade setting

SCORE has a number of concerns about the various indicators on which Ofqual proposes to judge performance and standards at GCSE and how this will impact the setting of grades.

### 6. Cohort changes and key stage 2 tests in grade setting

- SCORE is unconvinced by the claim that cohorts 'differ in ability year by year' and that 'if the cohort overall was particularly strong or weak compared with previous years... their GCSE performance might be expected to be stronger or weaker than that of previous cohorts.' SCORE would like to see the evidence on which this claim is based.
- SCORE asserts that there is no methodological system that can claim to set standards with absolute accuracy due to there being three overarching factors which affect the ability to analyse and interpret ability:
  - i. Comparison of standards across qualifications developed by multiple awarding organisations (AO) is highly problematic because examinations differ according to AO.
  - ii. Similarly, with several suites of qualifications on offer within the same subject it is highly difficult to compare standards across differing specifications.
  - iii. Given the range of qualifications on offer from AOs, it is extremely hard to set standards so that grades are balanced across subjects.
- With a cohort of around half a million students studying the sciences the distribution of abilities within the cohort is statistically indistinguishable from one year to the next, in the absence of external factors that can affect performance. With such a large sample size, and in the absence of external factors, the distribution of abilities within the cohort is constant. Variations in the apparent distribution

---

<sup>3 3</sup> See SCORE key stage 4 guidelines for SCORE's statement on guiding principles for assessment at key stage 4 (2013) <http://score-education.org/media/12525/ks4%20guidelines%20final%20version.pdf>; see also the SCORE principles for assessment of practical work for position on assessment specific to practical science (2014) <http://score-education.org/media/14286/score%20principles%20for%20the%20assessment%20of%20practical%20work%202014.pdf>

of abilities between such large cohorts, as measured through assessments, are an indication of systematic variations in the assessments themselves, and not of the distribution of abilities.

- In an environment in which no methodological system of standard setting can be entirely accurate, and in which cohort abilities do not change (external factors alter performance), SCORE recommends a norm-referenced approach to standard setting, in combination with other tools to support this approach (see paragraph 8). This would be no less applicable or accurate than the system that Ofqual is currently proposing.
- Ofqual proposes that key stage 2 data can provide a reliable set of information on which to predict the abilities of a GCSE cohort. However, testing and assessment can never be wholly reliable; this means that perceived variation in cohort standards at key stage 2 is much more likely to be a result of the variability of the test, not the cohort. Therefore basing standards on a measure of performance in a specific test at key stage 2 assumes an unrealistic reliability of that testing process and it is wholly unreasonable to use performance at that level as a calibrator at GCSE.

SCORE recommends that:

- Ofqual reconsider (or provide reliable, convincing evidence) to support the assertion that cohort abilities vary greatly on a year by year basis.

#### 7. A national reference test and grade setting

There are a number of issues surrounding the introduction by Ofqual of a national reference test in English language and mathematics, to be 'taken by a small cohort of Year 11 students shortly before they take their GCSEs' that 'will be representative of the national cohort'.

- Due to the small number of students involved in sitting a reference test, statistical variations between cohorts will automatically be introduced to results; such variations are entirely a reflection of the reference test and its small sample size rather than the ability of the cohort as a whole, which does not change year on year in the absence of external factors.
- Awarding organisations design examinations to different specifications and therefore there is no one assessment procedure on which to compare the reference test. Ofqual gives no explanation as to how a reference test could reliably feed in to assessment schemes that differ across multiple bodies.
- It is also unclear how within the short space of time between a reference test and GCSE examinations this information could be usefully conveyed in time for awarding organisations to use it.
- There is a concern that information about cohort performance taken from the English and mathematics tests would be used to feed into grade setting in other subjects. This would be inappropriate and unrepresentative of performance in other subject.
- It would be more valuable for Ofqual to focus all resources on the GCSE examinations themselves, rather than introduce an additional test that would require great effort to clarify and implement.

SCORE urges Ofqual to consider the following questions:

- a) How can Ofqual ensure that a reference test taken by a small cohort is representative of national standards across the cohort?
- b) How can information from the reference test be fed into standard setting of that year's GCSE examinations within such a short turnaround time?
- c) How would Ofqual ensure parity in the use of the reference test results across a multiple awarding organisation system?
- d) How can Ofqual ensure that the cohort sitting the national reference test is not placed under additional pressure during examination preparation time?

SCORE is also concerned whether or not b) and c) can be tackled in a robust way and therefore disagrees that the national reference should be used to feed into standard setting of GCSE examination results across subjects.

#### International standards and grade setting

- The Programme for International Students (PISA) tri-annual surveys run by the Organisation for Economic Co-operation and Development present a completely different set of data to that gained through GCSE qualifications, largely due to the two examinations being very different in purpose, content and process. Although international tests can give a useful comparison of educational systems across the world, Ofqual does not make clear how these international tests can assist in the setting of grade standards in GCSEs.
- Ofqual proposes to collect and review 'performance descriptions for qualifications and exams that serve a similar purpose to GCSEs in those countries whose students tend to perform well in international standards... [to] check whether or not... expectations about the standard required for a grade 5 is correct and inform users accordingly.'<sup>4</sup>

This proposal does not acknowledge that assessment designs and systems in other countries vary significantly from that of England's (not least because England's is one of very few education systems in the world operating a multiple awarding organisation structure), and that therefore comparability and transferability from one country's system to another would be extremely complex.<sup>5</sup>

- The suggestion also undermines the aim and purpose to ensure that England's GCSE assessment structure is first and foremost an educationally successful and robust system to begin with – rather than relying on external structures that may not be suitable in the English context.

#### SCORE recommends that:

- Ofqual reconsider making a link between PISA test data and standard setting in GCSEs.

#### 8. Norm-referencing

- No method of standard setting can claim to be entirely reliable in judging parity of grade standards year on year, particularly for subjects like the sciences where the cohort is large. Therefore SCORE believes it is important to make the system as simple, easily implemented and credible as possible, and that this should be the main aim of any standard setting approach.
- SCORE does not agree with Ofqual's conclusion that problems associated with norm-referencing rule out its application to standard setting (paragraphs 2.2.24-2.2.34). Given the large size of the national cohort studying GCSE sciences there is a very strong case for norm-referencing because the distribution of abilities within the cohort does not change significantly from one year to the next, in the absence of external factors. A system in which elements of norm-referencing are combined with additional checks and balances (for example, prior attainment of cohort; and independent monitoring data systems such as the Cem Centre's Advanced Level Information System) would achieve this by removing the need for an annual review of grade boundaries.
- In paragraph 3.3.32 of the consultation Ofqual states that 'We are considering establishing an expert committee to advise us how the evidence provided by the national reference test outcomes each

---

<sup>4</sup> Ofqual Consultation on Setting the Grade Standards of new GCSEs in England <http://comment.ofqual.gov.uk/setting-the-grade-standards-of-new-gcses-april-2014/> pp. 18-19

<sup>5</sup> For further information on differences of opinion towards the reliability of international comparison through PISA see: the Guardian 'OECD and Pisa tests are damaging education worldwide – academics', 6 May 2014 <http://www.theguardian.com/education/2014/may/06/oecd-pisa-tests-damaging-education-academics> and John Jerrim, 'England's "plummeting" PISA test scores between 2000 and 2009: Is the performance of our secondary school pupils really in relative decline?' [http://www.ioe.ac.uk/Study\\_Departments/J\\_Jerrim\\_qsswp1109.pdf](http://www.ioe.ac.uk/Study_Departments/J_Jerrim_qsswp1109.pdf)

year should be interpreted by the exam boards in their award of GCSEs.’ This plan demonstrates that ultimately a subjective approach to the use of evidence will be employed, thereby introducing all the unreliability and difficulty associated with anyone making decisions on how to feed statistical information into the grading system. The statement also introduces the question of how such an expert committee would be appointed and what role AOs would have in the process.

- Unreliability in any approach to standard setting is exacerbated in England by the existence of multiple AOs. The variation in qualifications on offer for each subject from across organisations introduces an additional level of complexity in standard setting as differing assessment procedures across examinations must be taken into account and this can only occur subjectively.

## 9. Standard setting across subjects

SCORE recommends that Ofqual take this opportunity to ensure that standards are set so that they are comparable across the full range of subjects offered at GCSE, and across the differing AO specifications. This would serve a number of purposes:

- a) It would ensure that the same grades in each subject accurately reflect the level expected in skills and knowledge (so that for example, an A in Music and an A in Biology reflect a similar level of demand). Evidence shows that this may not currently be the case.<sup>6</sup>
- b) Students can be deterred from pursuing the study of certain subjects at key stage 5 by the impression that those subjects are more challenging or severely graded. Where the common perception is that it is more difficult to achieve a high grade in a particular subject (such as in science or triple science), standard setting across subjects would go some way to reducing this inclination and offset the likelihood that students are deterred from further study in those subjects.
- c) As a result standard setting across the subjects would support students in making a more informed choice on subjects for post-16 study.

There is a strong body of research and evidence that compares methods by which cross subject standard setting could be achieved.<sup>7</sup>

### Introduction of two new grades in the highest achieving tier

1. Ofqual proposes to fit three grades (grade 7, 8 and 9) into the top award range (currently the A-A\* bracket)<sup>[1]</sup>. Within the current system, the percentage of marks required to achieve grades A and A\* varies across subjects and the boundaries for marking across qualifications can differ significantly. If the threshold for grade A\* is relatively low, say at 70%, a third grade in the top bracket may represent a useful tool for differentiation of high-performing candidates but, if the threshold is 85%, it will not. The root of this difficulty is the lack of comparability of standard of assessment between different qualifications and subjects, a longstanding issue that needs to be solved if any grading system is to have common currency.
2. SCORE suggests that a more positive approach to this problem would be solved through a two stage process. First, the marks for each qualification could be normalised using well-tested techniques

<sup>6</sup> In 2008, SCORE commissioned the CEM Centre, Durham University to conduct a full survey of research into the *Relative difficulty of examinations in different subjects*. For further evidence and discussion of the problems involved in and solutions available in setting comparability across subjects, see the full report <http://www.cem.org/attachments/SCORE2008report.pdf>

<sup>7</sup> See SCORE commissioned research report by CEM Centre, Durham University on the *Relative difficulty of examinations in different subjects* for a discussion of the different models and options for policy in cross subject standard setting: <http://www.cem.org/attachments/SCORE2008report.pdf>

<sup>[1]</sup> In 2013, a proportion of 21.3% achieved a grade A-A\*, while a proportion of 68.1% achieved within the range from A\*-C. Ofqual: *a brief explanation of summer 2013 GCSE results* <http://ofqual.gov.uk/news/summer-2013-gcse-results/>

based on prior attainment and subject pairs etc. Then, within each subject, grades are apportioned according to a norm-referenced standard setting system in which the percentage of the cohort in the top tier, or any grade, would be predefined.

### **Other grades**

It is well known that currently the grade C at GCSE is attached with great significance by schools, teachers, students and parents aiming to achieve 'pass' marks for students within the performance measures indicated by the accountability system. SCORE is concerned that by linking the new grade 4 with grade C the same impetus to 'pass' grade 4 will continue. While Ofqual acknowledges that lower grades, D,E,F and G 'can recognise real progress for some students'<sup>8</sup> SCORE recommends that Ofqual carefully consider what the meaning and purpose of the lower grades 1,2 and 3 are, in order to avoid undermining the progress of students at that level and pressuring schools towards grade 4.

### **Monitoring and maintaining standards**

3. A long-term plan for monitoring a system for standard setting should be embedded from the outset of the new proposed method. In the consultation document, paragraph 3.27 Ofqual explains how it will 'review and improve' grade descriptions without stipulating how it will review the process and system.<sup>9</sup>
4. SCORE does not agree that this demonstrates a robust or logical plan for monitoring and evaluating the changes and effects that the new standard setting procedure might have.

SCORE recommends that:

- Ofqual publish plans for how it will monitor and evaluate the new standard setting system.

---

<sup>8</sup> Ofqual Consultation on Setting the Grade Standards of new GCSEs in England <http://comment.ofqual.gov.uk/setting-the-grade-standards-of-new-gcses-april-2014/> p. 20

<sup>9</sup> Ofqual Consultation on Setting the Grade Standards of new GCSEs in England <http://comment.ofqual.gov.uk/setting-the-grade-standards-of-new-gcses-april-2014/> pp. 21